DOCUMENT RESUME

ED 040 029 RE 002 820

AUTHOR Powell, William R.
TITLE The Validity of the Instructional Reading Level.
PUB DATE 8 May 70
NOTE 20p.; Paper presented at the International Reading Association conference, Anaheim, Cal., May 6-9, 1970

EDRS PRICE EDRS Price MF-$0.25 HC-$1.10
DESCRIPTORS *Classroom Observation Techniques, *Informal Reading Inventory, Performance Criteria, Reading Comprehension, Reading Materials, Teacher Education, *Test Interpretation, *Test Validity, Word Recognition

ABSTRACT

Presented is a critical inquiry about the product of the informal reading inventory (IRI) and about some of the elements used in the process of determining that product. Recent developments on this topic are briefly reviewed. Questions are raised concerning what is a suitable criterion level for word recognition. The original criterion of 95 percent correct pronunciation for word recognition is considered too high. The application of one set of performance standards uniformly across all grade levels is questioned. Neither quantitative nor qualitative uniformity across passage levels is considered appropriate in dealing with errors. It is noted that present knowledge of the IRI precludes definitive statements concerning the hierarchial relation of the independent, instructional, and frustrational reading levels. The real value of the IRI is seen as affording the possibility of evaluating reading behavior in depth and as offering potential for training prospective teachers about reading behavior. References and tables are included. (WB)

EDO 40029

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

THE VALIDITY OF THE INSTRUCTIONAL READING LEVEL

William R. Powell
University of Illinois at Urbana-Champaign

What I would like to do is give you my proposition first, then go back through the process of analysis and development that brought me to these conclusions.

The real value of the informal reading inventory lies not so much in its identification of the instructional reading level, and by interpolation the independent and frustration levels; rather, its real value is that it affords the possibility of evaluating reading behavior in depth. Furthermore,

RE 002 820

Paper presented at the International Reading Association Convention, Anaheim, California, May 8, 1970.

it has the potential for training prospective teachers about reading behavior, a potential unequalled by other types of learning opportunities. For purposes of training teachers, the process becomes the product.

The strength of the IRI is not as a test instrument, but as a strategy for studying the behavior of the learner in a reading situation and as a basis for instant diagnosis in the teaching environment.

What we are really concerned with is the degree of mastery. The child does not have an *instructional* level; he has only a *performance* level. To obtain the desired performance level, adjustment has to be made in the criterion levels, the learning time, or the linguistic complexity of the written language. The selection of the adjustment variables is a teacher task, and therefore an instructional one.

When we speak of instructional level, we are referring to a teacher task; when we speak of performance, we are referring to the learner's behavior; and when we speak of difficulty of material, we are referring to the characteristics of the media. For maximum learning, all three have to match: performance level (child), instructional level (teacher); and passage difficulty (material). The instruction should be provided *by the teacher* at the performance level *of the child* that will allow for the exclusion of interfering or disruptive reading behaviors.

BACKGROUND

Statements and comments on the informal reading inventory are not new. Indeed, many papers on this general topic have been presented at conferences such as this one. But in the last few years, the nature of the discussion has shifted from one of description and exposition to one of inquiry and critical analysis. This altered perspective now is focusing on the critical issues--generating critical questions in an open forum about the concept, criteria, application, and empirical basis of the IRI, which has become a part of the fabric of reading instruction since its structured formulation by Betts (2) nearly thirty years ago.

A major product derived from the use of the IRI is the identification of three distinct reading levels--independent, instructional, and frustration. For instructional purposes, the assumption has been that each literate individual, regardless of maturity, has three such levels. Supposedly, these would be in hierarchical order in relationship to the difficulty of the materials, with the independent reading level being the lowest, or easiest, of the three. The other two levels, instructional and frustration, follow in ascending order as the readability of the material

increases. Each reading level is alleged to have specific instructional implications for the classroom teacher. While the existence of three different reading levels for literate persons is a powerful concept, it would have to be considered presently as a functionally useful but unvalidated construct.

Because the use of an IRI embodies most of the elements of the instructional environment, this process offers potential beyond the important task of making a match between children and suitable materials. There is the opportunity for teachers to gain diagnostic insights, from the simple indication of level to the complex evaluation of reading behavior. The latent power of this process is just beginning to be tapped as a means of expanding the conceptual framework of individuals in teacher education programs.

PURPOSE AND LIMITATIONS

Contrary to the possible implications of the title of the paper, I shall explore some of the facets and perceptions beyond the limited range of the instructional level. The fact that I do not expand broadly into other related dimensions of the IRI, I trust, will not be taken as a lack of sensitivity to the probable issues there. Components such as comprehension, rate, and symptoms of difficulty all play their interacting parts in affecting the total reading performance.

Rather than elaborate on the descriptive elements of the informal reading inventory, I am going to assume that you are

somewhat familiar with its characteristics, construction, and administration, as well as with at least one scoring scheme used for the interpretation of levels. These assumptions are made for the sake of expediency, so we can get on to the real purpose of the paper without undue delay. For those who wish to pursue information about the fundamental constitutents of the IRI, I would refer you to Betts (2), Johnson and Kress (10), and Zintz (18).

The purpose of this paper is to present a critical inquiry about the product of the informal reading inventory, and about some of the elements used in the process of determinirg that product. To achieve this purpose I propose to review recent developments on this topic briefly and to raise three particular questions. The first two deal with the process of the IRI, and the last with its product. These three questions are:

1. What is a suitable criterion level for word recognition in identifying the instructional reading level?

2. Is it appropriate to apply one set of performance standards uniformly across all grade levels?

3. Could it be that the major product of the IRI, i.e., the identification of three distinct reading levels, is a misinterpretation?

RECENT INQUIRIES

Without much doubt, the most widely used predetermined standards for evaluating reading performance on the IRI are those originally suggested by Betts (2). His criteria are:

Level	Word Recognition(%)	Comprehension(%)	Symptoms of Difficulty
Independent	99	90	none
Instructional	95	75	none
Frustration	90	50	some

Through the years, several individuals have expressed reservations and concern about the original criteria, but few have suggested other standards of performance that differed markedly.[1] In 1968, at the IRA convention in Boston, I broke the "silence of doubt" and openly challenged the existing sets of criteria (12). My investigation suggested that the original criteria simply are not consistent with the actual reading behavior of children. The Betts' criteria for the word-recognition dimension in evaluating oral reading behavior for the instructional reading level are too stringent, even for the proficient readers. The alternate set of criteria I found to be more consistent with children's actual performance are presented in Table I.

Only a year ago at the IRA convention in Kansas City, one full symposium program was devoted to the validity of the IRI. These presentations have subsequently been published (8). Particularily noteworthy out of that symposium collection was a paper by H. O. Beldin (1). He sytematically traced the

---

[1] N. B. Smith (15) is a noteable exception to this statement. Since 1959, she has proposed a lower percentage for correct pronunciation. Smith suggests an 80- to 85-percent accuracy range. Spache (16) has also offered an opinion that the Betts standards are arbitrarily too high.

historical development of the informal reading inventory and pointed out some of the issues regarding the process of this instrument.

Last November at the NCTE convention in Washington, D. C., Colin Dunkeld and I (13) presented further comparative data concerning the validity of the criteria I had suggested in the earlier paper. We compared sets of criteria from eight sources, five of which were derived from commonly used oral reading tests. This data is presented for your inspection in Table II. Attention should be called to the similarity of the criteria in the first four columns. Also, please note that only one of the word-recognition error ratios (on the Gilmore at the eighth grade) reached the predetermined standards originally set by Betts.

QUESTIONS AND ISSUES

What is a suitable criterion level for word recognition in identifying the instructional reading level? We have enough evidence to suggest what is an unsuitable criterion, but not enough yet to say with assurance what is suitable. It definitely would appear that the original criterion of 95-percent correct pronunciation (word-recognition)--that is, one error in every twenty running words--is too high for all age-grade levels.

The way two occurrences relate tends to support this conclusion. Studies have been conducted to evaluate other concurrent events using the original criteria, such as investigations comparing grade placement scores derived from standardized

reading measures with levels obtained from the informal reading inventory. In general, such studies have consistently indicated that scores from standardized tests vary at least from one to three years *above* a reported instructional reading level, as determined by the IRI.[1] While one study did clearly caution that generalizing from standardized scores to the instructional reading level was tenuous at best, a significant gap between the two types of assessment for a large number of the children studied did exist (6). Undoubtedly, the nature of the assessment process between the two types of instruments could be expected to produce a difference between scores. Nevertheless, the degree of difference has been viewed with some suspicion as being greater than what should be expected for proficient readers.

Now, suppose we apply this information to the model generally used for determining reading disability. The model typically used is the degree of difference between the subject's estimated capacity and actual reading achievement, as determined by scores from a standardized test. If the difference between capacity and achievement equals or exceeds a predetermined cut-off point, then the child is said to be disabled. If we apply the difference between standardized reading achievement

---

[1]The studies by Killgallon (11), Daniels (5), Williams (17), Sipay (14), Davis (6), and Brown (3) all support the contention that standard tests tend to overestimate the instructional level. All studies except the one by Sipay used the Betts criteria with slight modification. For example, Williams adjusted the minimum acceptance in comprehension at the instructional level from 75 to 70 percent. Sipay, however, used the criteria suggested by Cooper (4) (see Table II). Since these criteria are even more rigorous than those developed by Betts, the same pattern was found.

measures and the instructional reading level and then add the discrepancy between estimated capacity scores and the reading achievement scores, an interesting phenomenon occurs. For most children of average ability with at least average reading achievement scores, their instructional reading level is not likely to be within the acceptable lower limits of their estimated capacity. Suppose the estimated capacity and reading achievement were to match perfectly; even so, the difference between their reading performance, as estimated by standardized tests, and their instructional reading level, as measured by the IRI, would be great enough to cause the instructional reading level to be outside the usually acceptable limits of normal reading behavior. Is this at all suitable? If the criteria for determining the instructional reading level were representative of children's actual reading performance, would the discrepancies noted above diminish? It would seem logical to assume that for youngsters of average ability and achievement, the instructional level should be within the tolerable limits of their estimated capacity.

Is it appropriate to apply one set of performance standards uniformly across all grade levels? Here, we need to divide our attention between the quantitative and the qualitative. The quantitative dimension refers to the numerical count of the errors or miscues used in computing the percent correct figure, or the word-recognition error ratio. The qualitative aspect of the issue refers to the types of errors or miscues that are permitted for computational purposes.

The data in Table II would not support an assumption that the same quantitative ratio or percentage figure can apply uniformly across all grade levels. Apparently, there is a differential function in oral reading miscues from grade level to grade level.

My earlier investigation, resulting in new criteria, implied that the change in the word-recognition error ratio was due to the age/grade of the child. While the maturity of the reader certainly would be a factor in such a shift of error ratio, I now believe that the important factor is not the age/grade relationship but the difficulty level of the passage.

The implications of this were made only too clear to me by a written comment from one of my graduate students:

> If we now decide to use the criteria for passage levels rather than the child's level in school, is our decision to do so founded on the evidence in your study? For the average child reading grade, it won't make much difference, but what about the sixth grader referred to the clinic experiencing difficulty in reading. On which basis do we judge his performance, on say first and second grade passage? There is a big difference between 1/8 and 1/18.[1]

Nevertheless, all available data seems to indicate that there is an inverse relationship between the difficulty, or readability, level of a passage and the number of word-recognition errors tolerated by a reader. That is, the easier the material, the higher the percent of miscues that can be permitted by the reader while still maintaining an acceptable understanding level of

---

[1] Comment by Patricia Stoll, contained in an intraoffice memo to the author.

the material read. Conversely, the more complex the written language, the fewer the number of deviations that can be so tolerated and still realize an acceptable comprehension level.

The key word in such a discussion as the one in the preceding paragraph is *tolerate* (1). What is meant by tolerate? It is the level of error difficulty or deviation from the expected response that is not detrimental to total reading performance. The tolerance level allows for a compensation or adjustment of the reader within *his* range of functioning. As error intolerance increases, the material and instruction must be adjusted downward; and as error intolerance decreases the adjustment should increase.

Before leaving the quantitative dimension of this issue, I would like to offer a point of curiosity. What relationship exists, if any, between the percent of word recognition deviations and sentence length? As the material increases in complexity and difficulty, the sentence length will also increase. Is there an inverse relationship between sentence length and error tolerance? Or is deep structure or some other linguistic factor the important variable, not sentence length?

Qualitatively, uniformity across passage levels would not appear to be appropriate either. The types of errors that significantly affect a reader's tolerance level are not uniform from level to level. That is to say that the types of significant errors between an average second grader and an average sixth grader are different, and should be. This observation is

based on a doctoral study currently near completion at the University of Illinois by Colin Dunkeld (7). It also coincides with the types of findings by Yetta Goodman (9) in her study of oral reading miscues. She states, "It became evident that the type of miscues which beginning readers made change qualitatively as they become more proficient readers." Therefore, certain types of miscues in the reading of a passage of second grade difficulty might not be scoreable errors at that level, but might be used for determining error ratios at the fourth grade difficulty level, and vice-versa.

An apparent problem concerning the qualitative value placed on errors depends on the definition and classification used in processing those errors. There is little agreement among authorities on what constitutes a substitution, a mispronunciation, etc. The lack of agreement is not only in the basic definition, but also in the implications. Certainly, if error types are to have relevance and provide cues for instruction, then a reasonable degree of common interpretation will have to be established.

Could it be that the major product of the IRI, i.e., the identification of three distinct reading levels, is a misinterpretation? To search for truth, one has to be willing to risk the ultimate. To critically analyze the process and product of the IRI, one has to consider that the ultimate answer may be negative--that indeed the IRI has no actual validity, and that we who work with it are making something out of it that it is not. But that finding would offer positive direction for other types of options.

Research evidence to support the construct of an instructional reading level is minimal and incomplete; likewise, for the frustration reading level. This does not mean that we do not believe such levels exist. It simply means we do not yet have the data to support our beliefs.

One of the traditional beliefs regarding reading levels is that they form a hierarchical sequence--independent, instructional, and frustration, in that order. Spache challenges that opinion by reversing the position of the instructional and independent reading levels. He orders the levels this way: instructional, independent, and frustration.

There is absolutely no empirical data for defining the rank order nor the limits of the independent reading level. It has been assumed to be beyond the upper limits of the instructional level; therefore, Spache's reversal of the rank order may well be correct. How would we know which sequence is correct?

Since everyone is guessing about the location of the independent reading level, I might as well offer a conjecture on the subject. My impression is that the independent reading level is not static, but "floats." It may not always be located above or below the instructional reading level. The leverage to the reader is the interest value of the ideas and concepts. The greater the interest, the higher the passage difficulty can be for the independent reading level of a particular pupil. Conceivably, interest could cause this level to be quite variable, and it may be equal to or above the instructional level in specific types of materials. It is possible that for brief,

transitory, high-intensity periods, the interest value could project the independent reading level into the usual frustration zone (defined as beyond the lower limits of the instructional level). But until we have some data with which to define the limits of the independent level, your guess is as good as the three just given.

Another option may be that we are applying the right labels to the wrong agent. What we are really concerned with is the degree of mastery. The child does not have an *instructional* level; he only has a *performance* level. To obtain the desired performance level, adjustment has to be made in criterion levels, the learning time, or the linguistic complexity of the written language. The selection of the adjustment variables is a teacher task, and therefore an instructional one. When we speak of instructional level, we are referring to a teacher task; when we speak of performance, we are referring to the learner's behavior; and when we speak of difficulty of material, we are referring to the characteristics of the media.

For maximum learning, all three have to match: performance level (child), instructional level (teacher), and passage difficulty (material). The instruction should be provided by the teacher *at the performance level* of the child *that will allow for the exclusion of interfering or disruptive reading behaviors.*

CONCLUDING STATEMENT

The value of the IRI lies not in its identification of what has been called the instructional level (and the other levels by interpolation, because there are probably more effective and efficient methods of accomplishing such tasks. The use of cloze procedure is one alternative already available that has a considerable body of research data to support it.

The real value of the IRI is that it affords the possibility of *evaluating reading behavior in depth*. Furthermore, it has the potential for training prospective teachers about reading behavior, a potential unequalled by other types of learning opportunities. For purposes of training teachers, the process becomes the product.

The strength of the IRI is not as a test instrument, but as a strategy for studying the behavior of the learner in a reading situation and as a basis for instant diagnosis in the teaching environment.

BIBLIOGRAPHY

1. Beldin, H. O. "Informal Reading Testing: Historical Review and Review of the Research," pp. 67-84, in *Reading Difficulties: Diagnosis, Correction, and Remediation.* William K. Durr, editor. Newark, Del.: International Reading Association, 1970.

2. Betts, Emmett A. *Foundations of Reading Instruction.* New York: American Book Company, 1946.

3. Brown, Sandra R. "A Comparison of Five Widely Used Standardized Reading Test Scores and an Informal Reading Inventory for a Selected Group of Elementary Children." Unpublished Doctoral Dissertation, University of Georgia, 1963.

4. Cooper, J. Louis. "The Effect of Adjustment of Basal Reading Materials on Reading Achievement." Unpublished Doctoral Dissertation, Boston University, 1952.

5. Daniels, John E. "The Effectiveness of Various Procedures in Reading Level Placement," *Elementary English,* 39:590-600, October, 1962.

6. Davis, Sister M. Catherine Elizabeth. "The Relative Effectiveness of Certain Evaluative Criteria for Determining Reading Levels." Unpublished Doctoral Dissertation, Temple University, 1964.

7. Dunkeld, Colin G. "The Validity of the Informal Reading Inventory for the Designation of Instructional Reading Levels: A Study of the Relationships Between Children's Gains in Reading Achievement and the Difficulty of Instructional Materials." Doctoral Dissertation (in progress), University of Illinois, 1970.

8. Durr, William K. (editor). *Reading Difficulties: Diagnosis, Correction, and Remediation.* Newark, Del.: International Reading Association, 1970.

9. Goodman, Yetta. "Studies of Reading Miscues." From translated remarks made in Symposium II, *Applications of Psycholinguistics to Key Problems in Reading.* Kansas City, International Reading Association Convention, 1969.

10. Johnson, Marjorie S. and Kress, Roy A. *Informal Reading Inventories.* Reading Aid Series. Newark, Del.: International Reading Association, 1965.

11. Killgallon, Patsy A. "A Study of Relationships Among Certain Pupil Adjustments in Reading Situations." Unpublished Doctoral Dissertation, The Pennsylvania State University, 1942.

12. Powell, William R. "Reappraising the Criteria for Interpreting Informal Inventories," pp. 100-109, in *Reading Diagnosis and Evaluation*. Dorothy L. DeBoer, editor. Newark, Del.: International Reading Association, 1970.

13. Powell, William R. and Dunkeld, Colin G. "Validity of the IRI Reading Levels," *Elementary English* (in press).

14. Sipay, Edward R. "A Comparison of Standardized Reading Achievement Test Scores and Functional Reading Levels." Unpublished Doctoral Dissertation, University of Connecticut, 1961. (See also, *The Reading Teacher*, 17:265-268, January, 1964.)

15. Smith, Nila B. *Graded Selections for Informal Reading: Diagnosis for Grades 1 through 3*. New York: New York University Press, 1959.

16. Spache, George D. *Reading in the Elementary School*. Boston: Allyn and Bacon, Inc., 1964.

17. Williams, Joan. "A Comparison of Standardized Reading Test Scores and Informal Reading Inventory Scores." Unpublished Doctoral Dissertation, Southern Illinois University, 1963.

18. Zintz, Miles V. *The Reading Process: The Teacher and the Learner*. Dubuque, Iowa: Wm. C. Brown Company Publishers, 1970.

REFERENCES TO TESTS IN TABLE II

Durrell, Donald D. *Durrell Analysis of Reading Difficulty.* New York: Harcourt, Brace and World, Inc., 1955.

Gates, Arthur I. and McKillop, Anna S. *Gates-McKillop Reading Diagnostic Tests.* New York: Bureau of Publications, Teachers College, Columbia University, 1962.

Gilmore, John V. *Gilmore Oral Reading Test.* New York: Harcourt, Brace and World, Inc., 1968.

Gray, William S. *Gray Oral Reading Tests.* H. M. Robinson, editor. Indianapolis: The Bobbs-Merrill Co., 1963.

Spache, George D. *Diagnostic Reading Scales.* Monterey, California: California Test Bureau, 1963.

Table I

REVISED SCORING CRITERIA FOR THE INFORMAL READING INVENTORY (IRI)

Passages 1-2

WORD RECOGNITION

INDEPENDENT	INSTRUCTIONAL	FRUSTRATION
1/99-1/50	1/49-1/8	1/7 (AND BELOW)

COMPREHENSION

100%-90%	89%-70%	69% OR LESS

Passages 3-5

WORD RECOGNITION

INDEPENDENT	INSTRUCTIONAL	FRUSTRATION
1/99-1/50	1/49-1/13	1/12 (AND BELOW)

COMPREHENSION

100%-90%	89%-70%	69% OR LESS

Passages 6+

WORD RECOGNITION

INDEPENDENT	INSTRUCTIONAL	FRUSTRATION
1/99-1/50	1/49-1/18	1/17 (AND BELOW)

COMPREHENSION

100%-90%	89%-70%	69% OR LESS

William R. Powell
University of Illinois, Urbana
IRA, Anaheim, 1970

Table II

WORD RECOGNITION ERROR RATIOS
BY EIGHT SETS OF CRITERIA

Levels \ Criteria	Powell	Spache	Durrell	Gilmore	Gray	Gates McKillop	Betts-Killgallon	Cooper
P		1/4	1/3	1/3	1/7		1/20	1/50
1^2	1/6	1/5		1/5	1/8		1/20	1/50
2^1		1/8	1/8	1/6	1/11		1/20	1/50
2^2	1/8	1/7	1/9			1/2	1/20	1/50
3^1		1/10		1/8	1/11		1/20	1/50
3^2	1/11	1/13	1/12			1/3	1/20	1/50
4	1/13	1/15	1/13	1/11	1/10	1/4	1/20	1/25
5	1/12	1/16	1/16	1/13	1/11	1/6	1/20	1/25
6	1/17	1/16	1/18	1/14	1/9	1/6	1/20	1/25
7		1/16	1/17	1/18	1/10	1/6	1/20	
8		1/18		1/20	1/9	1/6	1/20	

WRP:ps
M/EE-131